

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 1 362 917 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**19.11.2003 Bulletin 2003/47**

(51) Int Cl.<sup>7</sup>: **C12N 15/62, C07K 19/00,  
C12N 15/87, C07K 14/16,  
C07K 14/245**

(21) Application number: **03252970.3**

(22) Date of filing: **13.05.2003**

(84) Designated Contracting States:  
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR  
HU IE IT LI LU MC NL PT RO SE SI SK TR**  
Designated Extension States:  
**AL LT LV MK**

(72) Inventor: **Hu, Luen  
Taipei 112 (TW)**

(30) Priority: **13.05.2002 US 144549**

(74) Representative: **Caldwell, Judith Margaret et al  
David Keltie Associates  
Fleet Place House  
2 Fleet Place  
London EC4M 7ET (GB)**

(71) Applicant: **Geneshuttle Biopharm, Inc.  
Taipei 112 (TW)**

(54) **A fusion protein for use as vector**

(57) The present invention provides a fusion protein comprising a fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and the homologue or the functional equivalent derivatives thereof and ii) a membrane translocation sequence or the functional equivalent peptides and/or de-

rivatives thereof. The fusion protein is used as a vector for nucleic acids delivery in vitro and particularly in vivo for gene therapy and the production of transgenic animal.

**EP 1 362 917 A2**

**Description****Background of the Invention**5 **Field of the Invention**

[0001] The present invention relates to a fusion protein referred as a novel vector for delivering molecules into cells.

10 **Description of the Prior Art**

[0002] The transfer of genetic material into cells in mammals is of increasing therapeutic and commercial importance. For instance, gene therapy procedures are used to correct acquired and inherited genetic defects, cancer, and viral infection. The ability to express artificial genes in humans facilitates the prevention and/or cure of many major human diseases, including many diseases that are not amenable to treatment by other therapies. However, biological membranes are natural barriers central to compartmentalization in living systems. Therefore, the polypeptides and oligonucleotides are generally considered to be of limited therapeutic value. Many studies have been conducted to overcome the problem of delivering such polypeptides and oligonucleotides.

[0003] Most of the initial work focused on the use of retroviral vectors to transform these cells. However, numerous difficulties with retroviruses have been reported. For example, it is hard to infect certain cell types. Retroviruses typically enter cells via receptors and if such receptors are not present in the cell, or are not present in large numbers, the infection is not possible or efficient.

[0004] Many researchers developed liposome systems for delivering the polypeptides and oligonucleotides into the cells. Liposomes are small membrane-enclosed spheres that have been formed with the appropriate DNA entrapped within it. However, this system also has inherent problems. It is difficult to control the size of the liposome and hence the uniformity of delivery to individual cells. Additionally, it is difficult to prevent leakage of the contents of the liposome and as with other techniques, there is difficulty in directing cell-type specificity.

[0005] Recently, several small regions of proteins called protein transduction domains (PTDs) have been developed to transport molecules into cells (Fischer *et al.*, Bioconjugate Chem., Vol. 12, No. 6, 2001). Such PTD can translocate the cell membrane freely in a way that is receptor-, or transporter- independent, non-saturable, and consumes no energy. The PTD can get across the barrier of the cell membrane within less than one hour. The fundamental requirements for the creation, isolation and utilization of TAT-fusion proteins to affect mammalian cells were described in Becker-Hapak *et al.* Methods 24, 247-256, 2001. In Steven *et al.*, a series of synthetic PTDs that strengthen the  $\alpha$ -helical content and optimize the placement of arginine residues were synthesized. Several PTD peptides possessed significantly enhanced protein transduction potential compared with TAT (Steven *et al.*, Cancer Research 61, 474-477, January 15, 2001). Furthermore, USP 6,090,619 described the preparation of a novel non-viral vector, which can bind to desired DNA to form a combination useful to transfect diseased mitochondria of human or animal cells. USP 6,339,139 provided a gene transfer system binding to a growth factor receptor, comprising a 4-element complex gene transfer system consisting of ligand oligopeptide/polycationic polypeptide/endosome release oligopeptide/exogenous DNA or 3-element complex consisting of ligand oligopeptide/polycationic polypeptide/ exogenous DNA.

[0006] The delivery vector systems in the art were confronted with one or more obstacles as described below. First, the immune responses are elicited by viral vector and cationic liposome when they are injected into living animals. Second, the extra-cellular fluids of cells, like blood stream, may dilute or eventually digest the gene cargo, which results in the lost of gene to be delivered. Third, the phospholipid bi-layer of cell membrane forms a natural barrier against the entrance of the nucleotide molecules. Therefore, the large molecules like DNA gene cannot cross the cell membrane in a freely (active or passive) transportation way to deliver the foreign DNA gene cargo into the cells. Fourth, the gel-like cytoplasm is rich in proteases and/or nucleases milieus in which the DNA gene cargo was degraded by the mechanism of endosomal trapping. Fifth, the vector-DNA cargo in the cytoplasm of cell meets the second barrier -- nuclear membrane. The DNA gene cargo should pass the nuclear membrane and release the DNA gene cargo in the nucleus where the gene can act. Sixth, even if the DNA gene cargo has ability to enter the cell nucleus, the delivery efficiency is low.

[0007] Based on the above obstacles, there is a need to develop a delivery system for effectively delivering a desired molecule into cells or nuclei.

55 **Summary of the Invention**

[0008] The invention relates to a fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and the homologue or the functional equivalent derivatives thereof and ii) a membrane translocation sequence or the functional equivalent peptides and/or derivatives thereof.

**Brief Description of the Drawings**

[0009]

5 Figure 1 shows the construction of gene that encodes the fusion protein of invention, rTAT-CspA, is represented by simplified flow-chart. The CspA gene is amplified from *E. coli* by PCR with primer 1 and 2, and gained the double stranded DNA/RNA binding activity by the insertion of annealed synthetic oligo 1 and 2 that encodes reversed form of TAT-PTD peptide sequence is introduced into the CspA gene and then the resulting of modified gene of the invention (rTAT-CspA) was subcloned into the pET28a expression vector.

10 Figure 2 shows the construction of gene that encodes the fusion protein of invention, (SPKR)4 - iTAT -CspA, is represented by simplified flow-chart. The CspA gene is amplified from *E. coli* by PCR with primer 1 and 2, and gained the double stranded DNA/RNA binding activity by the insertion of annealed synthetic oligo 3 and 4 that encodes DNA condensation sequence (SPKR)4 and the PTD sequence is further introduced into the modified CspA gene by PCR amplification with primer 3 and 4 and then the resulting of modified gene of the invention was subcloned into the pET28a expression vector.

15 Figure 3 shows that the induction and purification of the fusion protein is performed by the manufacturer's instruction and purified proteins are analyzed onto the 5%-15% SDS-PAGE and stained gel with coomassie blue; lane 1 is total cell lysate of pET28a/fusion protein; lane 2 is sonicated supernatant 1 of pET28a/fusion protein; lane 3 is sonicated supernatant 2 of pET28a/fusion protein; lane 4 is sonicated pellets of pET28a/fusion protein; lane 5 is flow through fraction after Nickel-bead affinity column chromatography; lane 6 is wash fraction containing 20mM imidazole in 1X binding buffer; lane 7 is elution 1 fraction containing; and lane 8 is elution 2 fraction containing 60mM imidazole in elution buffer; lane 9 is protein molecular weight marker purchased from Amersham-Pharmacia Company; lane 10 is purified iTat protein (0.25µg); and lane 11 is purified rTat protein (1.0µg).

20 Figure 4 shows the gel retardation assay. 0.5µg reporter plasmid pEGFP-N1 (purchased from Clontech.com.) was incubated with various amount of protein vector, "rTAT-CspA" or "iTAT-CspA", in the presence of 1x PBS buffer. After incubation at room temperature for 30 min, the DNA-protein complexes were separated onto 1% agarose gel (1x TBE) and stained gel with EtBr.

25 Figure 5 shows the green fluorescent protein expression from Hela cells that were transfected (1) without DNA plasmid (left panel), (2) rTAT-CspA-pEGFP-N1 complexes (middle), and (3) (SPKR)4-iTAT-CspA-pEGFP-N1. After 1 hours of incubation of DNA-protein complexes with cells, 0.45mL serum-plus DMEM broth was added and incubated at 37°C 6hrs and then 1mL Serum-plus DMEM broth was added, and the cells were further incubation at 37°C for 24hrs and examined the EGFP gene expression by the fluorescent microscope.

30 Figure 6 shows the green fluorescent protein expression from the fusion protein vector of the invention iTAT-CspA-pEGFP-N1 complexes transfected mouse embryos. The zona-removed mouse embryos were transfected either with 0.75µg of DNA plasmid pcDNA3.1 (left panel) or the pEGFP-N1 reporter DNA plasmid with protein vector of the invention, iTAT-CspA (right panel), respectively. After the time of intervals of incubation, the mouse embryo were examined the green fluorescent protein expression by the fluorescent microscope.

35 Figure 7 shows the frozen tissue section of liver, from the mouse treated with the complexes of the protein vector of the invention rTAT- or iTAT-CspA-pEGFP-N1 DNA complexes by i.p. injection. After 48hr, the mice were sacrificed and then examined the green fluorescent protein expression in the liver by confocal laser scanning microscope (Leica.com.).

40 Figure 8 shows the electrophoretic mobility shift assay (EMSA) being performed by using the protein of the invention, rTAT and human 5S rRNA.

**Detailed Description of the Invention**

45 [0010] The modern biotechnology is solely constructed on the basis of DNA gene delivery platform; for example, somatic cells gene therapy including cystic fibrosis, insulin-deficient diabetes and hemophilia A and B, etc. and transgenic animals. All of these research fields and commercial importance issues are required to ensure a more efficient,

economic and safe gene delivery vector so as to achieve the goal of human healthcare improvement.

[0011] The present invention provides a new fusion protein used as a vector. The vector of the invention is a non-viral vector which can bind to desired nucleic acids and effectively delivery them into any organism such as animal, cell line and embryo. It can be produced in large quantity with the least amount of cost and broadly applied in various fields such as transgenic animal and gene therapy.

#### Recombinant Fusion Protein of the Invention

[0012] The "fusion protein" as used herein refers to a fusion of a first amino acid sequence encoding a target polypeptide with a second or more amino acid sequence(s) defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of a target protein.

[0013] The invention provides a fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and the homologue or the functional equivalent derivatives thereof and ii) a membrane translocation sequence or the functional equivalent peptides and/or derivatives thereof.

#### Cold Shock Domain

[0014] According to the invention, any suitable cold shock domain and the homologue or functional equivalent derivatives thereof may be used in the fusion protein of the invention. The reason for choosing cold shock domain (CSD) is fully illustrated by its advantages: (1) the CSD is a universally conserved nucleic acid binding domain, prevalently existing in the pro- and eukaryotic kingdoms and as a result it contributes to minimizing the risk of immune response or even lower such risk to naught. Therefore, the CSD is the most proper candidate for the construction of gene delivery protein vector. Since the CSD does not cause large immune response of the host that received the gene delivery vector; (2) the intrinsic nucleic acid binding behavior of the CSD can be accommodated with the different Na<sup>+</sup>/K<sup>+</sup> ions distribution in the outside and inside of a cell. It means that the CSD very likely will not lose the gene cargo during the process of delivery and achieve the high efficiency of gene delivery; and (3) recently Lescar *et al.* has proposed that the envelope glycoprotein of Alphaviruses, flaviviruses which adopt  $\beta$ barrel-type structure to inserting the membrane of host cells and therefore, fusogenic peptide in-frame fusion with CSD may escape from the endosomal trapping or fusing cellular membrane directly. (Lescar, J., Roussel, A., Wein, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105, 137-148. and Kuhn, R.J. *et al.* *Cell* 108, 717-725; 2002.)

[0015] The CSD has already been structurally well characterized, which refers to as anti-parallel, 5 stranded  $\beta$ -barrel structure which is denoted to the OB-fold. Therefore, the OB-fold family having the structure similar to the CSD can be used in the fusion protein of the invention. Many research results have indicated that the CSD is able to bind single stranded nucleic acids. For example, the RNP-1 and RNP-2 RNA binding motif and the phenylalanine, lysine residues on the peripheral surface of cold shock domain are responsible for single stranded DNA, RNA binding (Bycroft *et al.*, *Cell*, Vol. 88, 235-242, January 24, 1997). Izumi *et al.* indicated that the Y box-binding protein-1 binds single-stranded nucleic acids (Nucleic Acids Research, 2001, Vol. 29, No. 5, 1200-1207). Preferably, the cold shock domain is selected from the group consisting of CspA, CspB, CspC, CspD, rpl S1 RNA binding domain, the eukaryotic Y-box proteins, DNA binding Protein B (DBPB), DBPA, EF-1, m RNP3, mRNP4, FRG Y1, and nuclease-sensitive-element-binding protein 1 (NSEP 1). More preferably, the cold shock domain is selected from the group consisting of CspA, rpl S1 RNA binding domain, human YB-1 and DNA binding Protein B.

[0016] It will be understood that the CSDs of the invention are not limited to the above illustrated species but also include homologous sequences obtained from any other sources. According to the above disclosures, persons skilled in the art can realize that any CSD homologue capable of binding nucleic acids also can be used in the fusion protein of the invention.

[0017] According to the invention, the functional equivalent derivatives of the CSD and the homologue can also be used in the fusion protein of the invention. The CSD or the homologue thereof can be modified to bind double stranded nucleic acids. A "functional equivalent derivative" of the CSD and the homologue thereof includes molecules recombinantly produced or chemically synthesized polypeptides which function in a manner similar to the reference molecule to achieve a desired result. Thus, a functional equivalent derivatives of the CSD and its homologue encompasses derivatives including any single or multiple amino acid additions, substitutions and/or deletions occurring internally or at the amino or carboxy termini thereof, and including any modifications such as the insertion of DNA binding domain or DNA condensation domain. It has also been demonstrated that the CSD can be modified to be able to bind the double stranded DNA or RNA. For example, Wang *et al.* found that the modified cold shock domains of CspA by domain swapped with the positive charge amino acids from its eukaryotic counterparts YB-1 protein can result in gaining double-stranded DNA binding activity (Wang *et al.*, *Mol. Microbiology*, 38(3): 526-534, 2000).

[0018] According to the invention, the CSD can be modified to obtain the derivatives having double-strand-RNA/

DNA binding activity, which can be accomplished by inserting into the CSD either DNA condensation domain or DNA binding domain. The DNA condensation domain can bind more DNA and/or condense DNA molecule to form nucleosome-like structure which is resistant to DNase digestion thereby less fusion protein should be used. According to the invention, any suitable DNA condensation or binding domain that can modify cold shock domain/ribosomal protein S1 RNA binding domain to be bound to double stranded DNA, can be used in the invention. For example, the suitable DNA condensation or binding sequence is the DNA condensation domain (SPKR) 3-4 that is derived from histone, high mobility group (HMG) proteins or the nuclear localization sequences (NLS) that are rich with positive charge amino acid of arginine, lysine, such as SV40 large T antigen, Myc, YB protein and other putative NLS.

#### 10 Membrane Translocation Sequence

[0019] According to the invention, the membrane translocation sequence of the fusion protein are those that possess the ability to transverse biological membranes efficiently in a process termed protein transduction, such as protein transduction domain (PTD) and membrane fusion sequence and their functional equivalent peptides or derivatives. It is known that the protein transduction does not occur through a classical reporter-, transporter- or endosome-mediated fashion. The membrane translocation sequence can rapidly and efficiently transduce molecules into cells. Examples of the membrane translocation sequence are shown in table below:

Sequence	Name(origin of sequence)
<b>Protein Transduction Domain</b>	
RQIKIWFQNRRMKWKK (SEQ ID NO: 1)	Pantp(43-58)(Penetratin®)
Kkwkmrrnqfwkvqr (SEQ ID NO: 2)	Retro-inverso pAntp(43-58)
RRWRRWWRRWWRRWRR (SEQ ID NO: 3)	W/R Penetratin
RRMKWKK (SEQ ID No: 4)	Pantp(52-58)
GRKKRRQRRRPPQ (SEQ ID NO: 5)	HIV TAT
YGRKKRRQRRR (SEQ ID No: 6)	HIV TAT
rrrrrrr (SEQ ID No: 7)	R7
Asp-Ala-Ala0Thr-Arg-Ser-Ala-Ala-Ser-Arg-Pro-Thr-Glu-Arg-Pro-Arg-Ala-Pro-Ala-Arg-Ser-Ala-Ser-Arg-Pro-Arg-Arg-Pro-Val-Glu(SEQ ID No: 8)	VP22(267~300)
<b>Membrane Fusion Sequence</b>	
GALFLGWLGAAGSTMGA (SEQ ID NO: 9)	Gp41 fusion sequence

5	GALFLGFLGAAGSTMGAWSQPKSKRKV (SEQ ID NO: 10)	MPG (gp41 fusion sequence-SV40NLS)
	MGLGLHLLVLAAALQGAWSQPKKKRKV (SEQ ID NO: 11)	Caiman crocodylus Ig(v) light chain-SV40NLS
	PLSSIFSRIGDP (SEQ ID NO: 12)	PreS2-TLM
10	FWRGDLVDFQV (SEQ ID NO: 13)	VP3 core protein
	KFTIVFPHNQKGNWKNVPSNYHYCP (SEQ ID NO: 14)	VSV-G peptide
	AKRARLSTSFNPVYPYEDES (SEQ ID NO: 15)	Ad fiber
15	GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO: 16)	Transportan
	RGGRLSYSRRRFSTSTGR (SEQ ID NO: 17)	SynB1
	AAVALLPAVLLALLAP (SEQ ID NO: 18)	MPS (Kaposi FGF signal sequence)
20	AAVLLPVLLAAP (SEQ ID NO: 19)	MPS (Kaposi FGF signal sequence)
	VTVLALGALAG.VGVG (SEQ ID NO: 20)	MPS (human integrin $\beta$ 3 signal sequence)
25	VAYISRGGVSTYYSDTVKGRFTRQKYNKRA (SEQ ID NO: 21)	P3
	KLALKLALKALKAAALKLA (SEQ ID NO: 22)	Model amphiphilic peptide
	WEAKLAKALAKALAKHLAKALAKALKACE A (SEQ ID NO: 23)	KALA

[0020] One preferred embodiment of the invention is to provide a fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and the homologue or the functional equivalent derivative thereof and ii) a membrane transduction domain. It is known in the art that the PTD can deliver the nucleic acids across cell membrane and nuclear membrane. By combining the CSD and PTD, the fusion protein of the invention can successfully deliver the desired nucleic acids to cytoplasm and cell nucleus. Studies have found a number of PTDs and their properties, features and efficacies (Fischer *et al*, Bioconjugate Chem., 2001, Vol. 12, No. 6, 825-841; Lindgren *et al*, Tips-March 2000, Vol. 21; Becker-Hapak *et al*, Methods 24, 247-256, 2001; and Ho *et al*, Cancer Research 61, 474-477, January 15, 2001). For example, Tat is a transcription-activating factor of 86-102 amino acids in length, depending on the viral strain, and is involved in the replication of HIV. The minimum Tat transduction domain is the basic residues 49-57. The VP22, a 38-kDa structural protein from herpes simplex virus-1, has the remarkable property of intercellular transport. The minimum VP22 transduction domain is the basic residues 267-300. The Antp is the Drosophila homeotic transcription factor, which is composed of three  $\alpha$ -helices with a  $\beta$ -turn between helices 2 and 3. The third  $\alpha$ -helix (residues 43-58) of the Antp is required for transduction.

[0021] Preferably, the PTD is selected from the group consisting of SEQ ID NOs. 1-8. More preferably, the PTD is selected from the group consisting of SEQ ID NOs: 1, 2, 4, 5, 6 and 7.

[0022] Another embodiment of the invention is to provide a fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and the homologue or functional equivalent derivative thereof. It is also known in the art that membrane fusion sequence can translocate the molecules to the cell membrane. By combining the CSD and membrane fusion sequence, the fusion protein of the invention can deliver the desired nucleic acids to cytoplasm. To deliver the desired nucleic acids to the cell nucleus, the fusion protein should further comprise a nuclear localization sequence. For example, one representative membrane fusion sequence is PreS2-TLM. The PreS2 is a permeable peptide whose translocation motif is referred to as PreS2-TLM. The PreS2-TLM corresponds to an amphipathic  $\alpha$ -helix between residues 41 and 52 of the PreS2 protein. The peptides appear to be able to penetrate a variety of cells including plant cells. Preferably, the membrane fusion sequence is selected from the group consisting of SEQ ID NO: 9-23. More preferably, the membrane fusion sequence is selected from the group consisting of 9, 10, 11, 12, 13, 17, 19, 20 and 21.

## Protein Purification Tagged Sequence

[0023] According to the invention, the fusion protein can further comprise the protein purification tagged sequence. According to the invention, the purification tagged sequence is only an optional element of the fusion protein, which is just for protein purification. The fusion protein of the invention can be purified by chemical purification techniques. For example, the Tat-PTD containing protein can be purified by heparin column (Hakansson *et al.*, Protein science (2001), 10:2138-2139. Furthermore, DNA column also can be used for the purification of the fusion protein of invention.

[0024] According to the invention, the protein purification tagged sequence can be N- or C- terminal fusion with gene delivery protein vector. Preferably, the protein purification tagged sequence is selected from the group consisting of HA, GST, His6 tag. More preferably, the protein purification tagged sequence is His6 tag.

Preparation of the Fusion Protein of the Invention

[0025] The fusion protein of the invention can be produced by culturing a host cell transformed with a recombinant vector under conditions allowing expression of said protein and isolating the protein thereby produced.

[0026] A DNA construct, i.e. recombinant DNA molecule, suitable for the expression of a fusion protein according to the invention may be used in the invention. For expression of a protein of the invention, the DNA construct is cloned into an expression vector that expresses a fusion protein according to the invention. The expression vector is, of course, chosen according to the nature of the host cell chosen for expression of the protein. Suitable such expression vectors are available commercially. Expression is preferably carried out in a prokaryotic host, more preferably a microbial host, especially *E. coli*, when a suitable expression vector is a prokaryotic expression vector such as a bacterial plasmid of pET system provided from Novagen. Co.

Application of the Fusion protein of the Invention

[0027] The invention provides a fusion protein as a delivery vector. The fusion protein of the invention has high efficiency of transfection *in vitro* and particularly *in vivo*, which can deliver gene into cells, embryos and living animals. The gene delivery performed by the fusion protein can be performed in species-specific, even individual-specific ways. Therefore, the fusion protein of the invention is an effective tool in the field of gene therapy and the production of transgenic animals.

[0028] The following Examples are offered by way of illustration and not by way of limitation.

**EXAMPLES**

## Example 1

**Cloning and construction of the Fusion Protein of the Invention**

[0029] A single colony of *E. coli* strain of DH5 $\alpha$  was picked up and then added to 100  $\mu$ l of sterile water. The resulting solution was boiled for 5 minutes. After brief centrifugation, aliquot of 10  $\mu$ l was taken and used as the DNA template for CspA gene amplification by polymerase chain reaction (PCR). The primer 1 and primer 2 were used to amplify CspA gene from *E. coli*. The sequence of primer 1 and 2 are listed below:

## Primer 1:

5'-gctagcATGTCCGGTAAAATGACTGGTATCGTAAA-3' (SEQ ID NO: 24)

## Primer 2:

5'-ctcgagATTACAGGCTGGTTACGTTA-3' (SEQ ID NO: 25)

(Small letters indicated Nhe I cutting site. Underlined small letters indicated Xho I cutting site)

[0030] After 25 cycles of PCR (annealing temperature was set at 55°C for 1 min), the PCR product was analyzed by 1.2% agarose gel electrophoresis and then purified by QIAquick PCR products Purification Kit (QIAGEN, Co.). The purified PCR products were cloned into pGEM-T vector (Promega Co.) and the screening of recombinant plasmid was instructed by manufacturer's technical manual and the plasmid pGEM (T)/CspA was obtained and then verified by DNA sequencing. Therefore, such cold shock protein A (CspA; Cold shock domain) further modified by inserting the DNA condensation or binding domain in the cold shock domain of the invention is used to bind double stranded nucleic acids.

Fusion Protein I: The simplest form of the fusion protein is comprised by the combination of NLS+/PTD from TAT and CSD, which is referred as rTAT (see Figure 1). The DNA binding/NLS+/PTD, reversed form of Tat peptide sequence (rTat), was encoded by the following synthetic oligonucleotides:

#### Oligo (1):

5'-taTGGGTCGCCGTCGTCAACGTCGTAAAAAGCGCCGTT-3'  
(SEQ ID NO: 26)

#### Oligo (2):

5'-ctagAACCGCGCTTTTTACGACGTTGACGACGGCGACCCA-  
3' (SEQ ID NO: 27)

Fusion Protein II: The second form of fusion protein is comprised by the combination of NLS+/PTD from TAT and DNA condensation sequence of (SPKR)4 and CSD, which is referred as (SPKR)4 - iTAT -CspA (see Figure 2). The DNA condensation sequence, (SPKR)4, was encoded by the following synthetic oligonucleotides:

#### Oligo(3):

5'-taTGGGCTCTCCTAAACGCTCTCCTAAACGCTCT  
CCTAAACGCTCT CCTAAACGTGGTT-3' (SEQ ID NO: 28)

#### Oligo(4) :

5'-ctagAACACGTTTAGGAGAGCGTTTAGGAGAGCG  
TTTAGGAGAG CGTTTAGGAGAGCCCA-3' (SEQ ID NO:  
29)

[0031] 100 pmol of each of the oligo(1) and oligo(2) or oligo(3) and oligo(4) were mixed separately and heated at 95°C for 5 minutes. The annealed double-stranded oligonucleotides were slowly cooled to 4°C and stored at - 20°C overnight. The 5'-end labeling of annealed double-stranded oligonucleotides were instructed by the manufacturer's recommendation with T4 polynucleotide kinase (purchased from Amersham Co.) and ATP. After 1 hour of incubation at 37°C, the reaction was stopped by heat inactivation and then extracted by phenol:chloroform:isoamylalcohol (v/v/v = 25:24:1; Applichem Co.). The annealed double-stranded oligonucleotides were precipitated by glycogen (Roche Co.) and cold absolute ethanol (Merck Co.).

[0032] The synthetic oligonucleotides encode reversed form of Tat peptide sequence (rTat) or DNA condensation



sequence (SPKR)4 were cloned into the Nde I, Nhe I-treated pGEM (T)/CspA respectively by T4 DNA ligase (New England Biolab. Com.) and transformed into *E. coli* strain DH5 $\alpha$  to obtain pGEM (T)/rTAT-CspA and pGEM (T)/(SPKR) 4-CspA DNA plasmid, respectively. For further constructing modified (SPKR)4-CspA with membrane translocation activity, we used primer 3 and 4, which were complemented with the partial coding sequence of CspA, and PTD coding sequence from TAT was incorporated into primer 4 and the DNA plasmid pGEM (T)/(SPKR) 4-CspA was used as PCR template. After 25 cycles of PCR, the PTD sequence was introduced into the pGEM (T)/(SPKR) 4-CspA by primer 3 and 4 with Taq enzyme (purchased from Takara. Co.). The sequence of PTD primer 3 and 4 are listed below:

**Primer 3:**

5'-NNNNNGGATCCAGAGAAGTGTACGAACACATC-3'

**Primer 4:**

5'-NNNNNGGATCCCGCAAGAAACGCCGT  
CAACGCCGCAGAGGATCTCTGGACGAAGGTCAGAAA.

(Underlined letters indicated Bam HI cutting site.)

**[0033]** The annealing temperature is set at 54°C for 1 minute. After 25 cycles of PCR, the resulting PCR products were purified by QIAquick PCR purification kit and subjected to Dpn I and Bam HI digestion. After the restriction enzyme digestion, DNA was separated onto 1% agarose gel (1X TAE) and purified DNA by QIAquick gel extraction kit (QIAGEN. Co.). The purified DNA was self-ligated and then transformed into *E. coli* strain DH5 $\alpha$  to obtain the pGEM (T)/(SPKR) 4-iTat-CspA plasmid.

**Construction of Expression Vector**

**[0034]** The fusion protein sequence encoded by the pGEM(T)/rTAT-CspA or (SPKR)4-iTat-CspA plasmids were subjected to Nde I and Xho I digestion and then separated onto 1% agarose gel and excised the DNA fragments containing the coding sequence of rTAT-CspA or (SPKR)4-iTat-CspA fusion protein were used as DNA insert, respectively. Expression vector pET28a (Novagen Co.) was treated with the same restriction enzymes Nde I, Xho I and then ligated with DNA insert rTAT-CspA or (SPKR)4-iTat-CspA fusion proteins coding sequence respectively and then transformed *E. coli* strain of DH5 $\alpha$  to obtain pET28a/rTAT-CspA or pET28a/(SPKR)4-iTat-CspA and then pET expression vectors were subjected to DNA sequencing. The sequencing verified pET28a/rTAT-CspA or pET28a/(SPKR)4-iTat-CspA plasmids were transformed into *E. coli* strain BL21(DE3) codon-plus cell (purchased from Stratagene. com) respectively for the expression of fusion proteins of the invention.

**Expression and Purification of Fusion Protein of the Invention**

**[0035]** *E. coli* strain BL21 (DE3) codon-plus cells harboring pET28a/ rTAT-CspA or pET28a/(SPKR)4-iTat-CspA plasmids were grown on LB (Kan+, Cm+) plate. A single colony was picked up and inoculated into 2ml LB broth (Kan+, Cm+) and then incubated at 37°C for 4 hours. 1 mL aliquot of bacterial culture was inoculated into 400mL terrific broth (Kan+, Cm+) for further incubation until the optical density OD600 reached 0.6-0.8. The final concentration of 1mM IPTG were added and then further incubated at 37°C for 5 to 7 hours. The IPTG-induced *E. coli* cells were harvested by centrifugation. The resulting pellets were stored at -70°C in a freezer for further protein purification.

**Purification of the Fusion Protein of the Invention**

**[0036]** The purification of the protein of invention was performed according to the manufacturer's instruction manual of pET system. The IPTG-induced expression *E. coli* host pellet was resuspended in 1x His Bonding buffer (500mM NaCl, 20mM imidazole and 20mM Tris-Cl, pH7.9) containing 8Murea, and then disrupted cells by sonication. After ultra-centrifugation, the sonicated supernatant fractions were subjected to nickel-ion affinity column and after washing column by 10 volumes of 1x Binding buffer (500mM NaCl, 25mM imidazole and 20mM Tris-Cl, pH7.9) containing 8Murea and 1x Washing buffer (500mM NaCl, 40mM imidazole, 20mM Tris-Cl, pH7.9) containing 8M urea. The His6

## Example 5

**Delivering Reporter DNA Plasmid pEGFP-N1 into Mouse with the Fusion protein of the Invention**

5 [0040] 20ug of DNA plasmid pEGFP-N1 was mixed with the fusion protein of the invention rTAT-CspA or (SPKR) 4-iTat-CspA in the 1x PBS buffer, incubation at RT°C for 15 min and then intraperitoneally injected into mouse respectively. 48 hrs later, mice were sacrificed and their livers were subjected to frozen tissue section and then examined EGFP protein expression by the confocal laser scanning microscope (Leica. Co.). Figure 7 shows that the fusion protein of the invention rTAT-CspA or (SPKR)4-iTat-CspA could successfully deliver the DNA plasmid pEGFP-N1 to the liver  
10 cells *in vivo*.

## Example 6

**RNA binding Capacity of Fusion protein of the Invention**

15 [0041] The PCR products of human 5s rDNA were cloned into the pGEM (T) vector. (Promega. Co. A gift from Dr. Elong Lin). The pGEM (T)/h5S rDNA plasmid was digested with Spe I and then separated onto 1% agarose gel (1X TAE). The linear pGEM (T)/h5S rDNA fragments were excised from the agarose gel and purified by QIAquick gel purification kit (QIAGEN. Co.). The human 5S rRNAs were synthesized by using the RiboMAX/T7 large-scale RNA production system (Promega. Co.) and the uniformly labeling of human 5S rRNA with UTP [ $\alpha$ -32P] (800ci/mmol, 10mCi/  
20 mL ; NEN. Co.) was performed by the manufacturer's instruction manual. The *in vitro* synthesized human 5S rRNA was separated onto 12% UREA-PAGE and then purified from the gel by the method of UV shadowing. After ammonia acetate and ethanol precipitation of human 5S rRNA was denatured by heating on the boiled water bath for 5 min and then slowly cooled down to room temperature and stored at -20°C overnight. 125ng of human 5S rRNA was mixed  
25 with various amount of the protein of invention of rTAT protein, incubation at RT °C for 30 min and then treated with or without RNase T1 (10U; Ambion. Co.) for further 10 min incubation at 37°C. The binding complexes were separated into 8% native PAGE(0.5X TBE) and then fixed the gel with 10% acetic of further acid and dry gel and autoradiography. As shown in figure 8, the RNase-resistant binding complexes are indicated by symbol "\*\*\*".

EP 1 362 917 A2

<212> PRT  
 <213> HIV TAT  
 5 <400> 5  
 Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln  
 1 5 10  
 10 <210> 6  
 <211> 11  
 <212> PRT  
 <213> HIV TAT  
 15 <400> 6  
 Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg  
 1 5 10  
 20 <210> 7  
 <211> 7  
 <212> PRT  
 <213> R7  
 25 <400> 7  
 Arg Arg Arg Arg Arg Arg Arg  
 1 5  
 30 <210> 8  
 <211> 30  
 <212> PRT  
 <213> VP22  
 35 <400> 8  
 Asp Ala Ala Thr Arg Ser Ala Ala Ser Arg Pro Thr Glu Arg Pro Arg  
 1 5 10 15  
 40 Ala Pro Ala Arg Ser Ala Ser Arg Pro Arg Arg Pro Val Glu  
 20 25 30  
 45 <210> 9  
 <211> 17  
 <212> PRT  
 <213> Gp41 fusion sequence  
 <400> 9  
 50 Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly  
 1 5 10 15  
 Ala  
 55

EP 1 362 917 A2

5 <210> 10  
 <211> 27  
 <212> PRT  
 <213> MPG (Gp41 fusion sequence-SV40NLS)  
 <400> 10  
 Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly  
 1 5 10 15  
 10 Ala Trp Ser Gln Pro Lys Ser Lys Arg Lys Val  
 20 25  
 15 <210> 11  
 <211> 27  
 <212> PRT  
 <213> Caiman crocodylus Ig(v) light chain-SV40NLS  
 20 <400> 11  
 Met Gly Leu Gly Leu His Leu Leu Val Leu Ala Ala Ala Leu Gln Gly  
 1 5 10 15  
 25 Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val  
 20 25  
 30 <210> 12  
 <211> 12  
 <212> PRT  
 <213> Pres2-TLM  
 <400> 12  
 35 Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro  
 1 5 10  
 40 <210> 13  
 <211> 20  
 <212> PRT  
 <213> Ad fiber  
 <400> 13  
 45 Ala Lys Arg Ala Arg Leu Ser Thr Ser Phe Asn Pro Val Tyr Pro Tyr  
 1 5 10 15  
 50 Glu Asp Glu Ser  
 20  
 55 <210> 14  
 <211> 27  
 <212> PRT  
 <213> Transportan

# EP 1 362 917 A2

<400> 14

5 Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu  
1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu  
20 25

10

<210> 15  
<211> 18  
<212> PRT  
<213> SynB1

15

<400> 15

Arg Gly Gly Arg Leu Ser Tyr Ser Arg Arg Arg Phe Ser Thr Ser Thr  
1 5 10 15

20

Gly Arg

25

<210> 16  
<211> 16  
<212> PRT  
<213> MPS (Kaposi FGF signal sequence)

<400> 16

30

Ala Ala Val Ala Leu Leu Pro Ala Val Leu Leu Ala Leu Leu Ala Pro  
1 5 10 15

35

<210> 17  
<211> 12  
<212> PRT  
<213> MPS (Kaposi FGF signal sequence)

<400> 17

40

Ala Ala Val Leu Leu Pro Val Leu Leu Ala Ala Pro  
1 5 10

45

<210> 18  
<211> 15  
<212> PRT  
<213> human integrin  $\beta$ 3 signal sequence

<400> 18

50

Val Thr Val Leu Ala Leu Gly Ala Leu Ala Gly Val Gly Val Gly  
1 5 10 15

55

<210> 19  
<211> 30  
<212> PRT

# EP 1 362 917 A2

<213> P3  
 <400> 19  
 5 Val Ala Tyr Ile Ser Arg Gly Gly Val Ser Thr Tyr Tyr Ser Asp Thr  
 1 5 10 15  
 Val Lys Gly Arg Phe Thr Arg Gln Lys Tyr Asn Lys Arg Ala  
 10 20 25 30  
 <210> 20  
 <211> 18  
 <212> PRT  
 15 <213> Model amphiphilic peptide  
 <400> 20  
 Lys Leu Ala Leu Lys Leu Ala Leu Lys Ala Leu Lys Ala Ala Leu Lys  
 20 1 5 10 15  
 Leu Ala  
 25 <210> 21  
 <211> 30  
 <212> PRT  
 <213> KALA  
 30 <400> 21  
 Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala Leu Ala Lys His  
 1 5 10 15  
 35 Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Ala Cys Glu Ala  
 20 25 30  
 40 <210> 22  
 <211> 35  
 <212> DNA  
 <213> Primer 1  
 <400> 22  
 45 gctagcatgt ccggtaaaat gactggtatc gtaaa 35  
 <210> 23  
 <211> 26  
 <212> DNA  
 50 <213> Primer 2  
 <400> 23  
 ctcgaqatta caggctggtt acgtta 26  
 55 <210> 24

# EP 1 362 917 A2

	<211> 38	
	<212> DNA	
5	<213> Oligo (1)	
	<400> 24	
	tatgggtcgc cgtcgtcaac gtcgtaaaaa gcgcggtt	38
10	<210> 25	
	<211> 40	
	<212> DNA	
	<213> Oligo (2)	
	<400> 25	
15	ctagaaccgc gcttttttacg acgttgacga cggcgaccca	40
20	<210> 26	
	<211> 59	
	<212> DNA	
	<213> Oligo (3)	
	<400> 26	
	tatgggtctt cctaaacgct ctccctaaacg ctctctctaaa cgctctccta aacgtggtt	59
25	<210> 27	
	<211> 61	
	<212> DNA	
	<213> Oligo (4)	
30	<400> 27	
	ctagaaccac gtttaggaga gcgttttagga gagcggttag gagagcggtt aggagagccc	60
	a	61
35	<210> 28	
	<211> 27	
	<212> DNA	
	<213> Primer 3	
40	<400> 28	
	ggatccagag aagtgtacga acacatc	27
45	<210> 29	
	<211> 57	
	<212> DNA	
	<213> Primer 4	
	<400> 29	
50	ggatcccgca agaaacgccg tcaacgccgc agaggatctc tggacgaagg tcagaaa	57
55		

## Claims

1. A fusion protein for delivery of a desired molecule into cells or nuclei, comprising i) a cold shock domain and a  
5 homologous or a functional equivalent derivative thereof and ii) a membrane translocation sequence or the functional  
equivalent peptides and/or derivatives thereof.
2. The fusion protein of claim 1, wherein the cold shock domain is selected from the group consisting of CspA, CspB,  
CspC, CspD, rpl S1 RNA binding domain, eukaryotic Y-box proteins, DNA binding Protein B (DBPB), DBPA, EFE-  
1, m RNP3, mRNP4, FRG Y1 and nuclease-sensitive-element-binding protein 1 (NSEP 1).  
10
3. The fusion protein of claim 1, wherein the cold shock domain is selected from the group consisting of CspA, rpl  
S1 RNA binding domain, human YB-1 and DNA binding Protein A & B and FRG Y1.
4. The fusion protein of claim 1, wherein the functional equivalent derivative of cold shock domain is modified by  
15 inserting into the cold shock domain with a DNA condensation domain or a DNA binding domain.
5. The fusion protein of claim 4, wherein the DNA condensation or binding domain is selected from the group con-  
sisting of DNA condensation domain (SPKR) 3-4 and the positive charge nuclear localization sequences (NLS+).
- 20 6. The fusion protein of claim 1, wherein the membrane translocation sequence is protein transduction domain (PTD)  
or membrane fusion sequence.
7. The fusion protein of claim 1, wherein the membrane translocation sequence is a protein transduction domain  
(PTD) selected from the group consisting of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, and 8.  
25
8. The fusion protein of claim 1, wherein the membrane translocation sequence is a protein transduction domain  
(PTD) selected from the group consisting of SEQ ID NOs: 1, 2, 4, 5, 6 and 7.
9. The fusion protein of claim 1, wherein the membrane fusion sequence is selected from the group consisting of  
30 SEQ ID NOs: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, and 23.
10. The fusion protein of claim 1, wherein the membrane fusion sequence is selected from the group consisting of  
SEQ ID NOs: 9, 10, 11, 12, 13, 17, 19, 20 and 21.
- 35 11. The fusion protein of claim 1, which further comprises a protein purification tagged sequence.
12. The fusion protein of claim 11, wherein the protein purification tagged sequence is selected from the group con-  
sisting of HA, GST, and His6 tag.
- 40 13. The fusion protein of claim 1, for use in delivering nucleic acids selected from the group consisting of DNAs and  
RNAs to cells for gene delivery.
14. The fusion protein of claim 1, for use in delivering nucleic acids to *in vivo* cells for gene delivery.
- 45 15. The fusion protein of claim 1, for use in delivering nucleic acids to an embryo or to a living animal for the production  
of transgenic animal.
16. A fusion protein comprising i) a cold shock domain and the homologous or the functional equivalent derivatives  
thereof and ii) a protein transduction domain or the functional equivalent peptides and derivatives thereof.  
50
17. The fusion protein of claim 16, which further comprises protein purification tagged sequence.
18. The fusion protein of claim 17, wherein the protein purification tagged sequence is selected from the group con-  
sisting of HA, GST, and His6 tag.  
55
19. The fusion protein of claim 16, for use in delivering nucleic acids to cells for gene delivery.
20. The fusion protein of claim 16, for use in delivering nucleic acids to *in vivo* cells for gene delivery.



21. The fusion protein of claim 16, for use in delivering nucleic acids to an embryo or to a living animal for the production of transgenic animal.
- 5 22. A fusion protein comprising i) a cold shock domain and the homologue or the functional equivalent derivatives thereof and ii) a membrane fusion sequence or the functional equivalent peptides and derivatives thereof.
23. The fusion protein of claim 22, which further comprises a protein purification tagged sequence.
- 10 24. The fusion protein of claim 23, wherein the protein purification tagged sequence is selected from the group consisting of HA, GST, and His6 tag.
25. The fusion protein of claim 22, for use in delivering nucleic acids to cells for gene delivery.
- 15 26. The fusion protein of claim 22, for use in delivering nucleic acids to *in vivo* cells for gene delivery.
27. The fusion protein of claim 22, for use in delivering nucleic acids to an embryo or to a living animal for the production of transgenic animal.
- 20 28. The fusion protein of claim 22, which further comprises a nuclear localization sequence.
29. The fusion protein of claim 28, which further comprises a protein purification tagged sequence.
30. The fusion protein of claim 29, wherein the protein purification tagged sequence is selected from the group consisting of HA, GST, and His6 tag.
- 25 31. The fusion protein of claim 28, for use in delivering nucleic acids to cells for gene delivery.
32. The fusion protein of claim 28, for use in delivering nucleic acids to *in vivo* cells for gene delivery.
- 30 33. The fusion protein of claim 28, for use in delivering nucleic acids to an embryo or to a living animal for the production of transgenic animal.

35

40

45

50

55

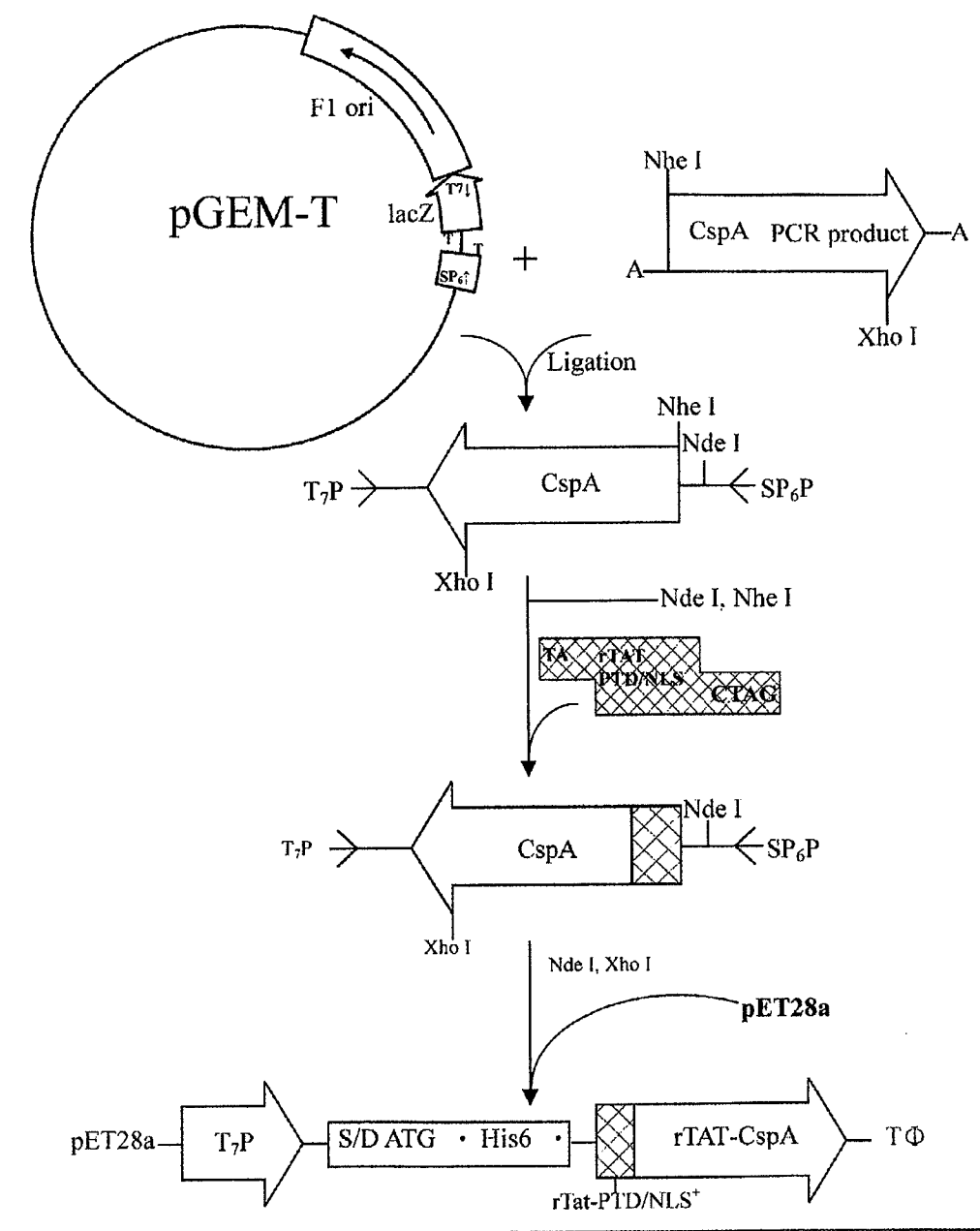


Fig.1.

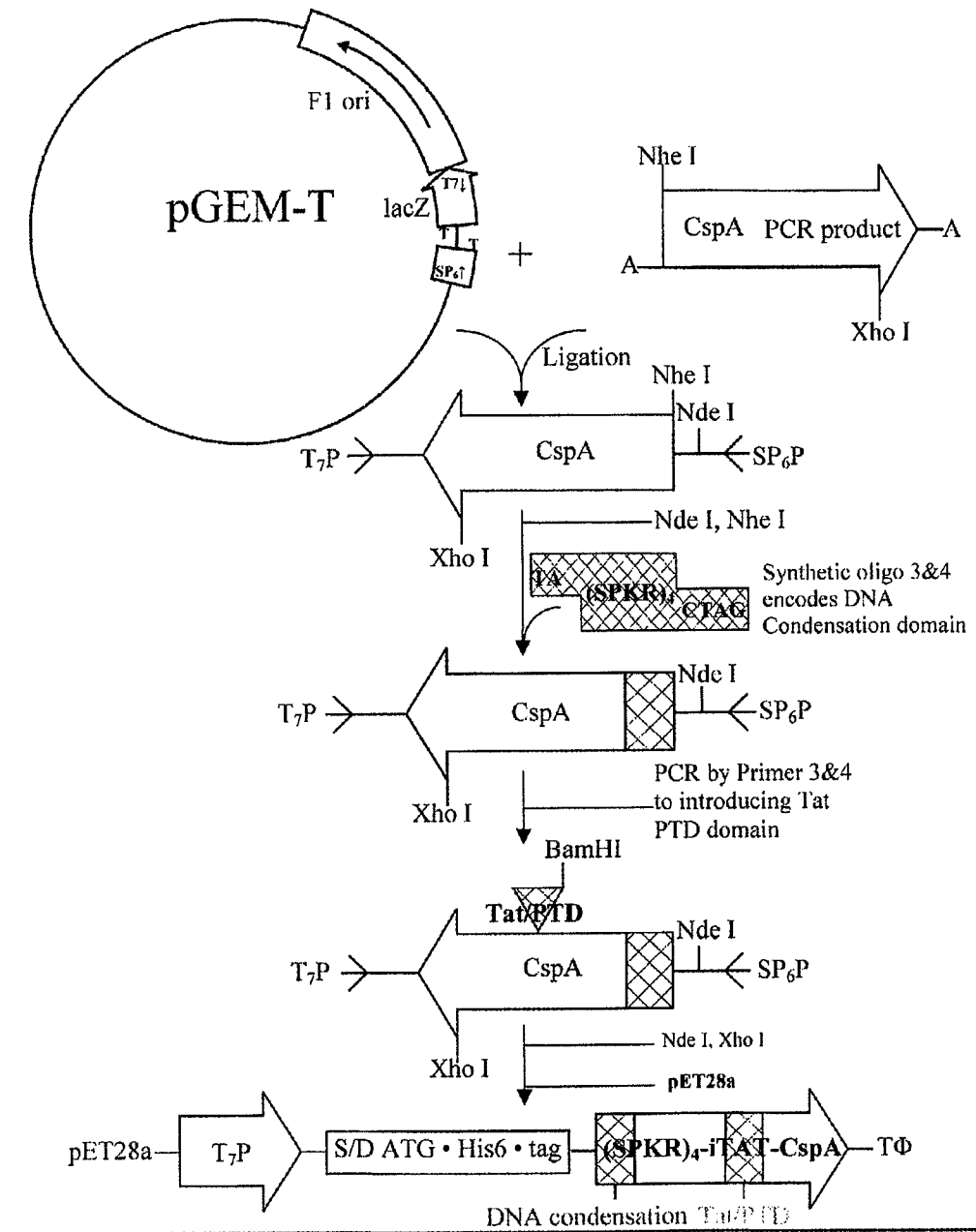


Fig.2.

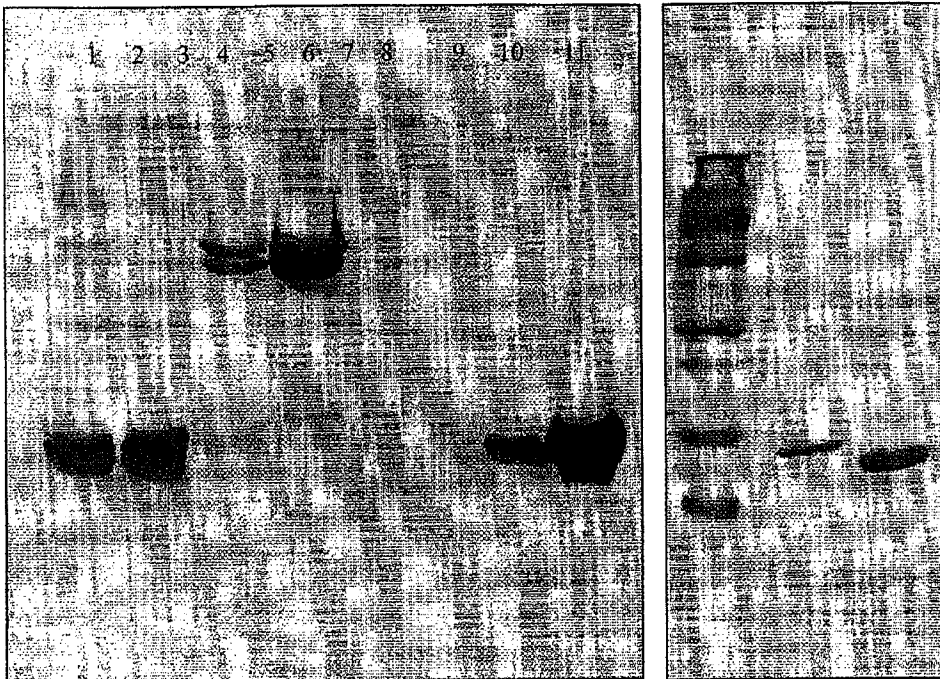


Fig.3.

B.

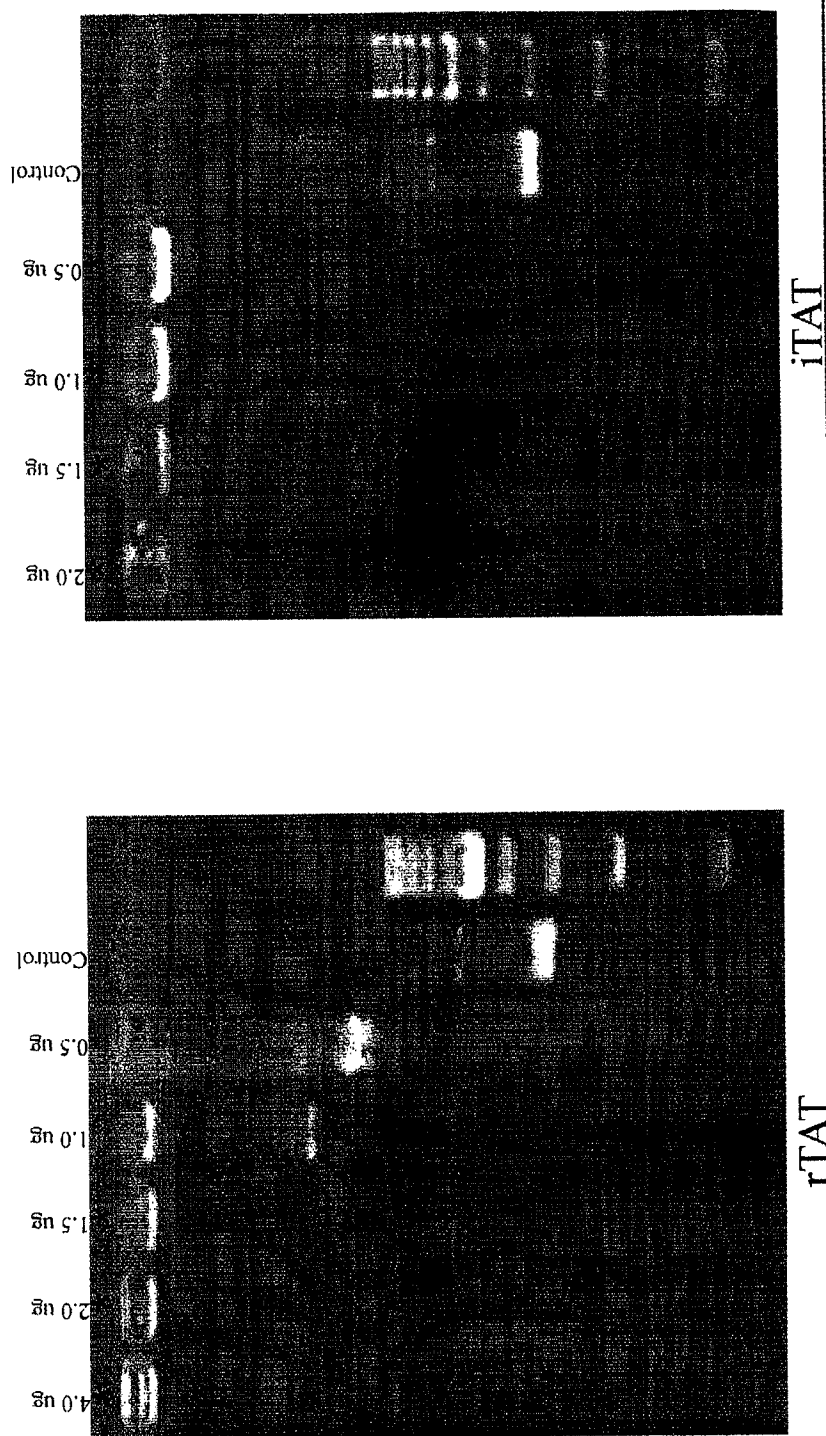


Fig.4.

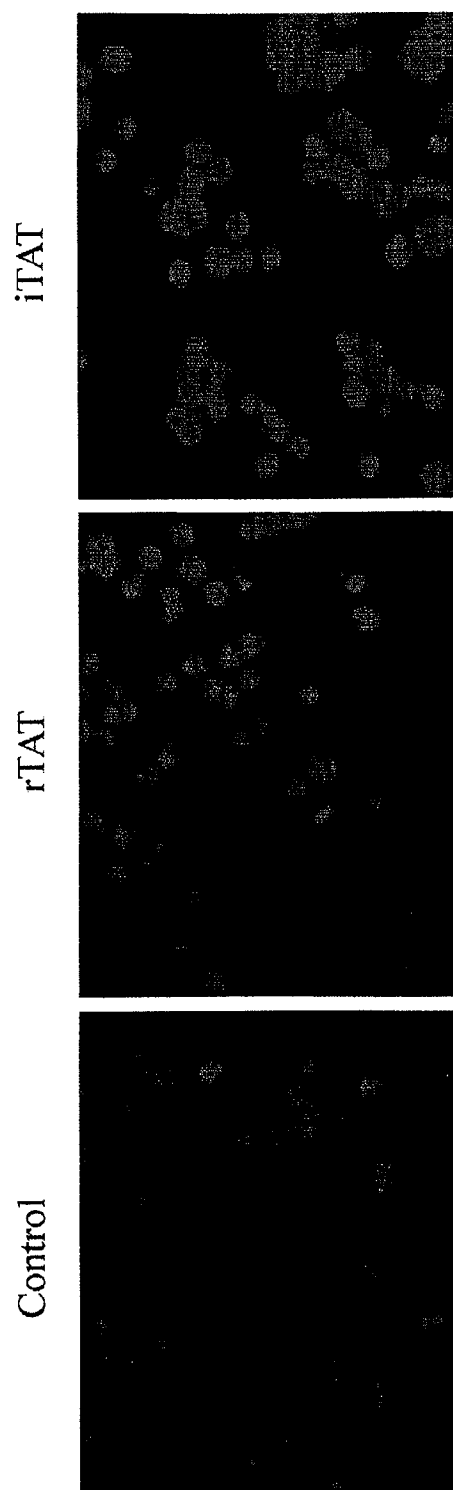


Fig. 5

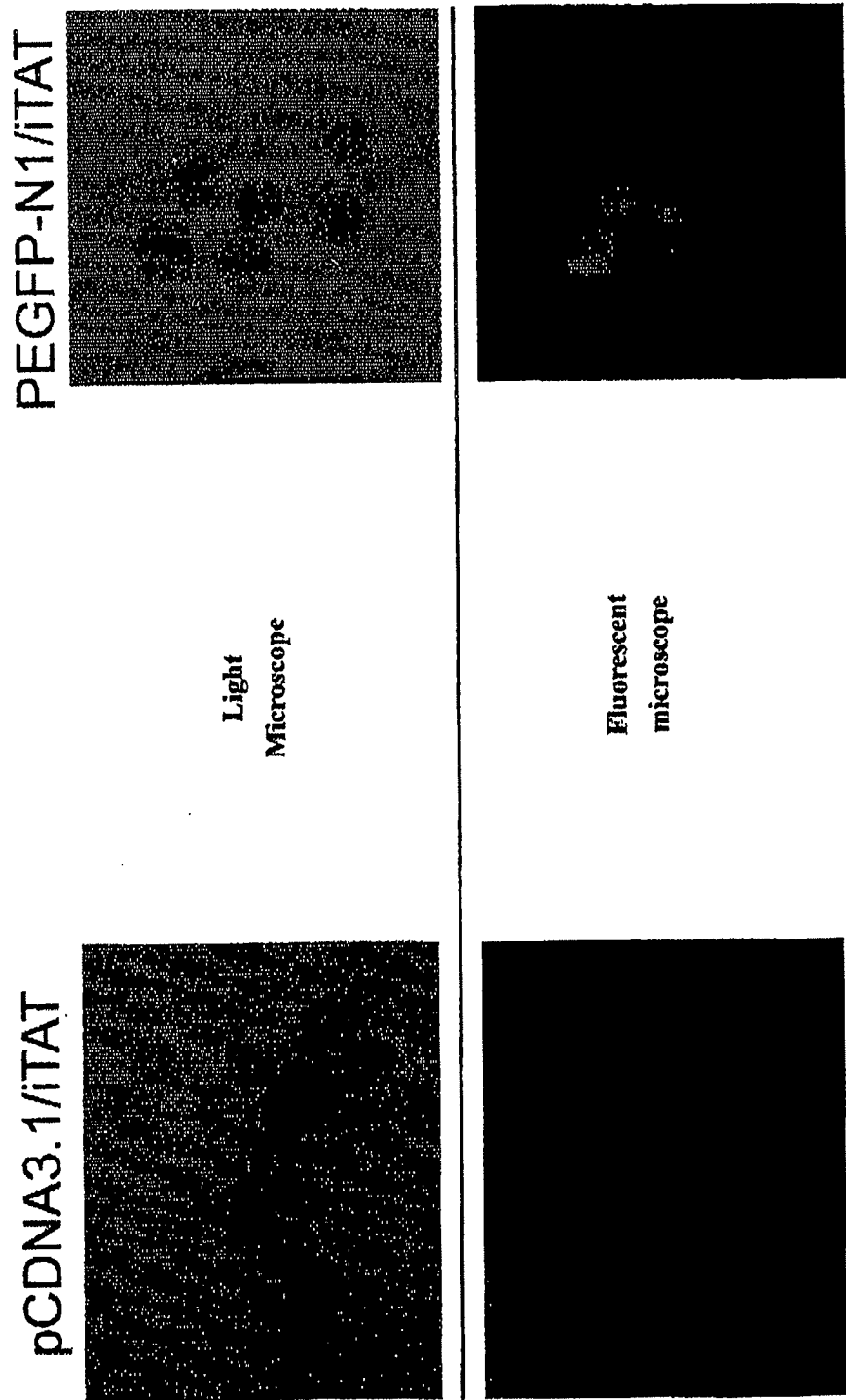


Fig. 6.

# Confocal microscope/Liver

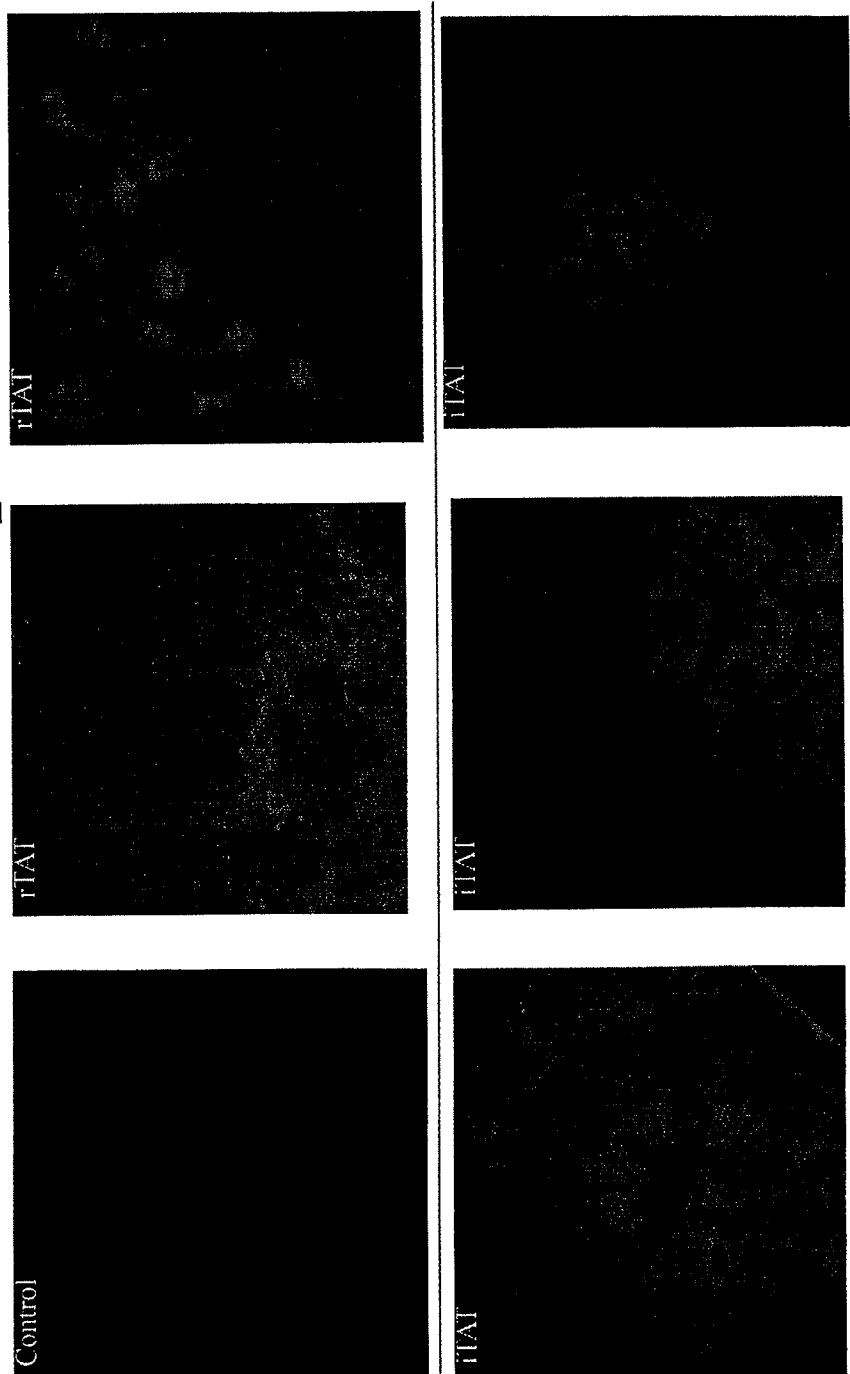


Fig. 7



	rTAT-CspA(ug)					h55rRNA probe					
	0.125	0.25	0.5	1.0	2.0	+	+	2.0	1.0	0.5	0.25
+	+	+	+	+	+	+	+	+	+	+	+
RNase T1(10U)	+	+	+	+	+	+	+	+	+	+	+

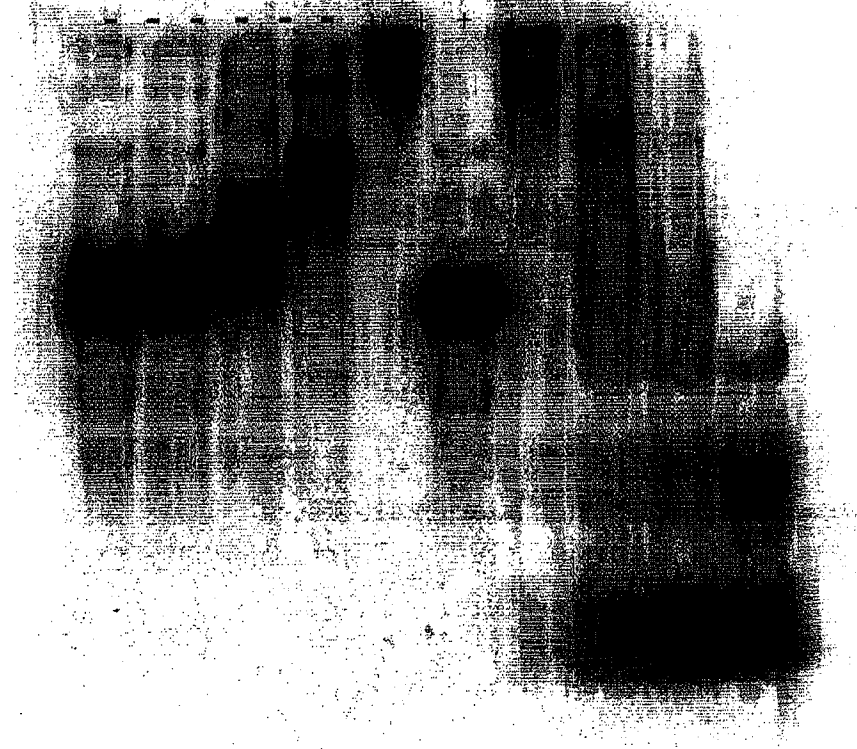


Fig.8.